sciendo This is an open access article licensed under the Creative Commons BY-NC-ND License.

Journal of Artificial General Intelligence 12(1) 1–25, 2021
DOI: 10.2478/jagi-2021-0001

Submitted 2020-10-16
Accepted 2021-01-08

Measuring Intelligence and Growth Rate: Variations on Hibbard's Intelligence Measure

Samuel Alexander SAMUELALLENALEXANDER@GMAIL.COM
(Primary Author)
The U.S. Securities and Exchange Commission, New York Regional Office

Bill Hibbard
Space Science and Engineering Center, University of Wisconsin, Madison, WI

Editor: Marcus Hutter

Abstract

In 2011, Hibbard suggested an intelligence measure for agents who compete in an adversarial sequence prediction game. We argue that Hibbard's idea should actually be considered as two separate ideas: first, that the intelligence of such agents can be measured based on the growth rates of the runtimes of the competitors that they defeat; and second, one specific (somewhat arbitrary) method for measuring said growth rates. Whereas Hibbard's intelligence measure is based on the latter growth-rate-measuring method, we survey other methods for measuring function growth rates, and exhibit the resulting Hibbard-like intelligence measures and taxonomies. Of particular interest, we obtain intelligence taxonomies based on Big-O and Big-Theta notation systems, which taxonomies are novel in that they challenge conventional notions of what an intelligence measure should look like. We discuss how intelligence measurement of sequence predictors can indirectly serve as intelligence measurement for agents with Artificial General Intelligence (AGIs).

1. Introduction

In his insightful paper, Hibbard (2011) introduces a novel intelligence measure (which we will here refer to as the *original Hibbard measure*) for agents who play a game of adversarial sequence prediction (Hibbard, 2008) "against a hierarchy of increasingly difficult sets of" evaders (environments that attempt to emit 1s and 0s in such a way as to evade prediction). The levels of Hibbard's hierarchy are labelled by natural numbers, and an agent's original Hibbard measure is the maximum $n \in \mathbb{N}$ such that said agent learns to predict all the evaders in the nth level of the hierarchy, or implicitly[1] an agent's original Hibbard measure is ∞ if said agent learns to predict all the evaders in all levels of Hibbard's hierarchy.

The hierarchy which Hibbard uses to measure intelligence is based on the growth rates of the runtimes of evaders. We will argue that Hibbard's idea is really a combination of two orthogonal ideas. First: that in some sense the intelligence of a predicting agent can be measured based on the growth rates of the runtimes of the evaders whom that predictor

1. Hibbard does not explicitly include the ∞ case in his definition, but in his Proposition 3 he refers to agents having "finite intelligence", and it is clear from context that by this he means agents who fail to predict some evader somewhere in the hierarchy.

learns to predict. Second: Hibbard proposed one particular method for measuring said growth rates. The growth rate measurement which Hibbard proposed yields a corresponding intelligence measure for these agents. We will argue that *any* method for measuring growth rates of functions yields a corresponding *adversarial sequence prediction intelligence* measure (or *ASPI* measure for short) provided the underlying number system provides a way of choosing canonical bounds for bounded sets. If the underlying number system does not provide a way of choosing canonical bounds for bounded sets, the growth-rate-measure will yield a corresponding ASPI taxonomy (like the big-O taxonomy of asymptotic complexity).

The particular method which Hibbard used to measure function growth rates is not very standard. We will survey other ways of measuring function growth rates, and these will yield corresponding ASPI measures and taxonomies.

The structure of the paper is as follows.

- In Section 2, we review the original Hibbard measure.
- In Section 3, we argue that any method of measuring growth rates of functions yields an ASPI measure or taxonomy, and that the original Hibbard measure is just a special case resulting from one particular method of measuring function growth rate.
- In Section 4, we consider Big-O notation and Big-Θ notation and define corresponding ASPI taxonomies.
- In Section 5, we consider solutions to the problem of measuring growth rates of functions using majorization hierarchies, and define corresponding ASPI measures.
- In Section 6, we consider solutions to the problem of measuring growth rates of functions using more abstract number systems, namely the hyperreal numbers and the surreal numbers. We do not assume previous familiarity with these number systems.
- In Section 7, we give pros and cons of different ASPI measures and taxonomies.
- In Section 8, we summarize and make concluding remarks.

2. Hibbard's original measure

Hibbard proposed an intelligence measure for measuring the intelligence of agents who compete to predict evaders in a game of adversarial sequence prediction (we define this formally below). A predictor p (whose intelligence we want to measure) competes against evaders e. In each step of the game, both predictor and evader simultaneously choose a binary digit, 1 or 0. Only after both of them have made their choice do they see which choice the other one made, and then the game proceeds to the next step. The predictor's goal in each round is to choose the same digit that the evader will choose; the evader's goal is to choose a different digit than the predictor. The predictor wins the game (and is said to *learn to predict* e, or simply to *learn* e) if, after finitely many initial steps, eventually the predictor always chooses the same digit as the evader.

Definition 1 *By B, we mean the binary alphabet $\{0,1\}$. By B^*, we mean the set of all finite binary sequences. By $\langle\rangle$ we mean the empty binary sequence.*

Definition 2 *(Predictors and evaders)*

1. *By a* predictor*, we mean a Turing machine*[2] *p which takes as input a finite (possibly empty) binary sequence $(x_1, \ldots, x_n) \in B^*$ (thought of as a sequence of* evasions*) and outputs* 0 *or* 1 *(thought of as a* prediction*), which output we write as $p(x_1, \ldots, x_n)$.*

2. *By an* evader*, we mean a Turing machine e which takes as input a finite (possibly empty) binary sequence $(y_1, \ldots, y_n) \in B^*$ (thought of as a sequence of* predictions*) and outputs* 0 *or* 1 *(thought of as an* evasion*), which output we write as $e(y_1, \ldots, y_n)$.*

3. *For any predictor p and evader e, the* result of p playing the game of adversarial sequence prediction against e *(or more simply, the* result of p playing against e*) is the infinite binary sequence $(x_1, y_1, x_2, y_2, \ldots)$ defined as follows:*

   (a) *The first evasion $x_1 = e(\langle\rangle)$ is the output of e when run on the empty prediction-sequence.*

   (b) *The first prediction $y_1 = p(\langle\rangle)$ is the output of p when run on the empty evasion-sequence.*

   (c) *For all $n > 0$, the $(n+1)$th evasion $x_{n+1} = e(y_1, \ldots, y_n)$ is the output of e on the sequence of the first n predictions.*

   (d) *For all $n > 0$, the $(n+1)$th prediction $y_{n+1} = p(x_1, \ldots, x_n)$ is the output of p on the sequence of the first n evasions.*

4. *Suppose $r = (x_1, y_1, x_2, y_2, \ldots)$ is the result of a predictor p playing against an evader e. For every $n \geq 1$, we say* the predictor wins round n in r *if $x_n = y_n$; otherwise,* the evader wins round n in r*. We say that p* learns to predict e *(or simply that p* learns e*) if there is some $N \in \mathbb{N}$ such that for all $n > N$, p is the winner of round n in r.*

Note that if e simply ignores its inputs $(y_1, \ldots, y_n)$ and instead computes $e(y_1, \ldots, y_n)$ based only on n, then e is essentially a sequence. Thus Definition 2 is a generalization of sequence prediction, which many authors have written about (such as Legg (2006), who gives many references). In the future, it could be interesting to consider variations of the game involving probability in various ways, for example, where the predictor wins if his guesses have $> 50\%$ win rate, or where the predictor states how confident he is about each guess, or other such variations.

In the following definition, we differ from Hibbard's original paper because of a minor (and fortunately, easy-to-fix) error there.

Definition 3 *Suppose e is an evader. For each $n \in \mathbb{N}$, let $t_e(n)$ be the maximum number of steps that e takes to run on any length-n sequence of binary digits. In other words, $t_e(0)$ is the number of steps e takes to run on $\langle\rangle$, and for all $n > 0$,*

$$t_e(n) = \max_{b_1,\ldots,b_n \in \{0,1\}} (\textit{number of steps } e \textit{ takes to run on } (b_1, \ldots, b_n)).$$

2. The measures we introduce in this paper would also work if we defined predictors as not-necessarily-computable functions $B^* \to B$, but this would not add much insight. We prefer to emphasize the duality between predictors and evaders when each is a Turing machine.

Example 4 *Let e be an evader. Then $t_e(2)$ is equal to the number of steps e takes to run on input $(0,0)$, or to run on input $(0,1)$, or to run on input $(1,0)$, or to run on input $(1,1)$—whichever of these four possibilities is largest.*

Definition 5 *Suppose $f : \mathbb{N} \to \mathbb{N}$ and $g : \mathbb{N} \to \mathbb{N}$. We say f* majorizes *g, written $f \succ g$, if there is some $n_0 \in \mathbb{N}$ such that for all $n > n_0$, $f(n) > g(n)$.*

Definition 6 *Suppose $f : \mathbb{N} \to \mathbb{N}$. We define E_f to be the set of all evaders e such that $f \succ t_e$.*

Definition 7 *(The original Hibbard measure) Let $g_1, g_2, \ldots$ be the enumeration of the primitive recursive functions given by Liu (1960). For each $m > 0$, define $f_m : \mathbb{N} \to \mathbb{N}$ by*

$$f_m(k) = \max_{0<i\leq m} \max_{j\leq k} g_i(j).$$

For any predictor p, we define the original Hibbard intelligence *of p to be the maximum $m > 0$ such that p learns to predict e for every $e \in E_{f_m}$ (or 0 if there is no such m, or ∞ if p learns to predict e for every $e \in E_{f_m}$ for every $m > 0$).*

The following result shows that the measure in Definition 7 does not overshoot the agents being measured.

Proposition 8 *For any integer $m \geq 1$, there is a predictor p with original Hibbard measure $\geq m$.*

Proof This is part of Proposition 3 of Hibbard (2011), but we give a self-contained proof here because we will state similar results about other measures below.

For each computable $f : \mathbb{N} \to \mathbb{N}$, let p_f be the predictor who proceeds as follows when given evasion-sequence $(x_1, \ldots, x_n) \in B^*$ as input.

First, by calling itself recursively on inputs

$$\langle\rangle, (x_1), (x_1, x_2), \ldots, (x_1, \ldots, x_{n-1}),$$

p_f determines the prediction-sequence $(y_1, \ldots, y_n)$ as in Definition 2.

Next, p_f considers the first n Turing machines, $T_1, \ldots, T_n$. For $1 \leq i \leq n$, say that T_i is an *nth-order e-lookalike* if the following requirements hold:

- If T_i halts in $\leq f(0)$ steps on input $\langle\rangle$, then T_i outputs x_1 on that input.
- If T_i halts in $\leq f(1)$ steps on input (y_1), then T_i outputs x_2 on that input.
- If T_i halts in $\leq f(2)$ steps on input (y_1, y_2), then T_i outputs x_3 on that input.
- ...
- If T_i halts in $\leq f(n-1)$ steps on input $(y_1, \ldots, y_{n-1})$, then T_i outputs x_n on that input.
- T_i halts in $\leq f(n)$ steps on input $(y_1, \ldots, y_n)$ with some output $X_{i,n}$.

By simulating $T_1, \ldots, T_n$ as needed (which only requires finitely many steps), p_f determines if any T_i is an nth-order e-lookalike. If so, p_f outputs $p_f(x_1, \ldots, x_n) = X_{i,n}$ for the minimal such i. If not, p_f outputs 0.

Claim: For every computable $f : \mathbb{N} \to \mathbb{N}$, for every evader e, if $(x_1, y_1, x_2, y_2, \ldots)$ is the result of p_f playing against e, then for all-but-finitely-many $n \in \mathbb{N}$, if $f(n) > t_e(n)$ then $x_{n+1} = y_{n+1}$.

Let $f, e, (x_1, y_1, \ldots)$ be as in the claim and assume $f(n) > t_e(n)$. Being an evader, e is a Turing machine, say, the kth Turing machine. Since $f(n) > t_e(n)$, it follows that T_k is an nth-order e-lookalike. It follows that, on input $(x_1, \ldots, x_n)$, p_f will play the output given by some nth-order e-lookalike $T_{k'}$, $k' \leq k$.

For any $k' < k$, say that p_f *is tricked by* $T_{k'}$ *on input* $(x_1, \ldots, x_n)$ if, on said input, p identifies $T_{k'}$ as the first nth-order e-lookalike and so plays $y_{n+1} = X_{k',n}$ but $X_{k',n} \neq x_{n+1}$ (loosely speaking: p_f is led to believe the evader is $T_{k'}$, and this false belief causes p_f to incorrectly predict the evader's next digit). It follows that p_f will *not* identify $T_{k'}$ as an n'th-order e-lookalike when run on $(x_1, \ldots, x_{n'})$ for any $n' > n$. Thus, p_f can only be tricked at most once by $T_{k'}$ for any particular $k' < k$. If p_f is not so tricked, then either p_f identifies $T_{k'}$ as the first nth-order e-lookalike for some $k' < k$ (in which case p_f plays $y_{n+1} = X_{k',n} = x_{n+1}$, lest p_f be tricked by $T_{k'}$), or else p_f identifies T_k as the first nth-order e-lookalike, in which case p_f plays $y_{n+1} = X_{k,n} = x_{n+1}$ since $e = T_k$. Either way, after possibly finitely many exceptions caused by being tricked, p_f always plays $y_{n+1} = x_{n+1}$ whenever $f(n) > t_e(n)$, proving the claim.

We claim p_{f_m} has original Hibbard measure $\geq m$. To see this, let $e \in E_{f_m}$, we must show p learns e. Let $(x_1, y_1, \ldots)$ be the result of p_{f_m} playing against e. Since $e \in E_{f_m}$, $f_m \succ t_e$, so for-all-but-finitely-many $n \in \mathbb{N}$, $f_m(n) > t_e(n)$. And, by the above Claim, with at most finitely many exceptions, whenever $f_m(n) > t_e(n)$, $x_{n+1} = y_{n+1}$. It follows that p_{f_m} learns e, as desired. ■

Unfortunately, the original Hibbard measure is not computable (unless the background model of computation is contrived), as the following proposition shows[3].

Proposition 9 *Assume the background model of computation is well-behaved enough that there is an evader e_0 which always outputs 0 and whose runtime t_{e_0} is bounded by some primitive recursive function. Then the original Hibbard measure is not computable: there is no effectively computable procedure, given a predictor, to compute its original Hibbard measure. In fact, there is not even an effectively computable procedure to tell if one given predictor has a higher original Hibbard measure than another given predictor.*

Proof Let p_1 be a predictor which always outputs 1, and let m be its original Hibbard measure. By the existence of e_0, it follows that $m < \infty$ since certainly p_1 does not learn e_0. Let p be a predictor with original Hibbard measure $> m$ (Proposition 8).

If the proposition were false, then we could solve the halting problem as follows. Given any Turing machine M, in order to determine whether or not M halts, proceed as follows.

3. In fact, assuming a non-contrived background model of computation, for any strictly increasing total computable function f, even the following can be shown to be non-computable: given an evader e, to determine whether or not $f \succ t_e$.

Let p_M be the predictor which, on input $(x_1, \ldots, x_n)$, outputs $p(x_1, \ldots, x_n)$ if M halts in $\leq n$ steps, or 1 otherwise. Clearly, M halts if and only if p_M has a higher original Hibbard measure than p_1. ■

Likewise, the variations on Hibbard's measure which we present in this paper are also non-computable. To quantify their precise degrees of computability (e.g., where they fall within the arithmetical hierarchy) would be beyond the scope of this paper. We will, however, state one conjecture. If we modified the original Hibbard measure by replacing Liu's enumeration of the primitive recursive functions by an enumeration of all total computable functions, then we conjecture the resulting measure would be strictly higher in the arithmetical hierarchy (i.e., would require strictly stronger oracles to compute), essentially because the set of total computable functions is *not* computably enumerable, whereas the primitive recursive functions are.

2.1 Predictor intelligence and AGI intelligence

Definition 7, and similar measures and taxonomies which we will define later, quantify the intelligence of predictors in the game of adversarial sequence prediction. But any method for quantifying the intelligence of such predictors can also approximately quantify the intelligence of (suitably idealized) agents with Artificial General Intellience (that is, the intelligence of AGIs).

The idealized AGIs we have in mind should be capable of understanding, and obedient in following or trying to follow, commands issued in everyday human language[4] (this is not to say that all AGIs must necessarily be obedient, merely that for the purposes of this paper we restrict our attention to obedient AGIs). For example, if such an idealized AGI were commanded, "until further notice, compute and list the digits of pi," it would be capable of understanding that command, and would obediently compute said digits until commanded otherwise[5].

It is unclear how an AGI ought to respond if given an impossible command, such as "write a computer program that solves the halting problem", or Yampolskiy's "Disobey!" (Yampolskiy, 2020). But an AGI should be capable of understanding and attempting to obey an open-ended command, provided it is not impossible. For example, we could command an AGI to "until further notice, write an endless poem about trees," and the AGI should be able to do so, writing said poem line-by-line until we tell it to stop. This is despite the fact that the command is open-ended and under-determined (there are many decisions involved in writing a poem about trees, and we have left all these decisions to the AGI's discretion). The AGI's ability to obey such open-ended and under-determined commands exemplifies its ability to "adapt with insufficient knowledge and resources" (Wang, 2019). One well-known example of an open-ended command which an AGI should be perfectly

4. It is somewhat unclear how explicitly an AGI would obey certain commands. To use an example of Yampolskiy (2020), if we asked a car-driving AGI to stop the car, would the AGI stop the car in the middle of traffic, or would it pull over to the side first? We assume this ambiguity does not apply when we ask the AGI to perform tasks of a sufficiently abstract and mathematical nature.
5. Our thinking here is reminiscent of some remarks of Yampolskiy (2013).

capable of attempting to obey (perhaps at peril to us) is Bostrom's "manufacture as many paperclips as possible" (Bostrom, 2003).

In particular, such an idealized AGI X should be capable of obeying the following command: "Act as a predictor in the game of adversarial sequence prediction". By giving X this command, and then immediately filtering out all X's sensory input except only for input about the digits chosen by an evader, we would obtain a formal predictor in the sense of Definition 2. This predictor might be called "the predictor generated by X". Strictly speaking, if the command is given to X at time t, then it would be more proper to call the resulting predictor "the predictor generated by X at time t": up until time t, the observations X makes about the universe might have an effect on the strategy X chooses to take once commanded to act as a predictor; but as long as we filter X's sensory input immediately after giving X the command, no further such observations can so alter X's strategy. In short, to use Yampolskiy's terminology (Yampolskiy, 2012), the act of trying to predict adversarial sequence evaders is *AI-easy*.

Thus, any intelligence measure (or taxonomy) for predictors also serves as an intelligence measure (or taxonomy) for suitably idealized AGIs. Namely: the intelligence level of an AGI X is equal to the intelligence level of X's predictor. Of course, a priori, X might be very intelligent at various other things while being poor at sequence prediction, or vice versa, so this only approximately captures X's true intelligence.

Of course, the same could be said for any competency measure on any task: we do not make any claims that when we measure X's intelligence via X's predictor's performance, that this is in any sense "the one true intelligence measure". One could just as well measure X's intelligence in terms of the Elo ranking X would obtain if one ordered X to compete at chess. We would offer two motivations to consider adversarial sequence prediction ability as a particularly interesting proxy for AGI intelligence measurement:

1. There seem to be high-level connections between intelligence and prediction in general (Hutter, 2004), of which adversarial sequence prediction is an elegant and parsimonious example.

2. Adversarial sequence prediction ability is not bounded above, in the sense that for any particular predictor p, one can easily produce a predictor p^+ that learns all the evaders which p learns and at least one additional evader.

3. Quantifying growth rates of functions

The following is a general and open-ended problem.

Problem 10 *Quantify the growth-rate of functions from* $\mathbb{N}$ *to* $\mathbb{N}$.

The definition of the original Hibbard measure (Definition 7) can be thought of as implicitly depending on a specific solution to Problem 10, which we make explicit in the following definition.

Definition 11 *For each* $m > 0$*, let* f_m *be as in Definition 7. For each* $f : \mathbb{N} \to \mathbb{N}$*, we define the* original Hibbard growth rate $H(f)$ *to be* $\min\{m > 0 : f_m \succ f\}$ *if there is any such* $m > 0$*, and otherwise* $H(f) = \infty$.

In order to generalize the original Hibbard definition in a uniform way, we will rearrange notation somewhat. We have, in some sense, more notation than necessary, namely the notation in (Hibbard, 2011) and synonymous notation which is modified to generalize more readily.

Lemma 12 *For every natural $m > 0$ and every $f : \mathbb{N} \to \mathbb{N}$, $H(f) \leq m$ if and only if $f_m \succ f$.*

Proof Straightforward. ■

Definition 13 *For every $m \in \mathbb{N}$, let E_m^H be the set of all evaders e such that $H(t_e) \leq m$.*

Lemma 14 *For every natural $m > 0$, $E_m^H = E_{f_m}$.*

Proof Let e be an evader. By Definition 13, $e \in E_m^H$ if and only if $H(t_e) \leq m$. By Lemma 12, $H(t_e) \leq m$ if and only if $f_m \succ t_e$. But by Definition 6, this is the case if and only if $e \in E_{f_m}$. ■

Corollary 15 *For every predictor p, the original Hibbard measure of p is equal to the maximum natural $m > 0$ such that p learns e whenever $e \in E_m^H$, or is equal to 0 if there is no such m, or is equal to ∞ if p learns e whenever $e \in E_m^H$ for all $m > 0$.*

Proof Immediate by Lemma 14 and Definition 7. ■

In other words, if S is the set of all the m as in Corollary 15, then the original Hibbard measure of p is the "canonical upper bound" of S, where by the "canonical upper bound" of a set of natural numbers we mean the maximum element of that set (or ∞ if that set is unbounded).

Remark 16 *Corollary 15 shows that the definition of the original Hibbard measure can be rephrased in such a way as to show that it depends in a uniform way on a particular solution to Problem 10, namely on the solution proposed by Definition 11. For* any *solution H' to Problem 10, we could define evader-sets $E_m^{H'}$ in a similar way to Definition 13, and, by copying Corollary 15, we could obtain a corresponding intelligence measure given by H' (provided there be some way of choosing canonical bounds of bounded sets in the underlying number system—if not, we would have to be content with a taxonomy rather than a measure, a predictor's intelligence falling into many nested taxa corresponding to many different upper bounds, just as in Big-O notation a function can simultaneously be $O(n^2)$ and $O(n^3)$). This formalizes what we claimed in the Introduction, that Hibbard's idea can be decomposed into two sub-ideas, firstly, that a predictor's intelligence can be classified in terms of the growth rates of the runtimes of the evaders it learns, and secondly, a particular method (Definition 11) of measuring those growth rates (i.e., a particular solution to Problem 10).*

3.1 A theoretical note on the difficulty of Problem 10

In this subsection, we will argue that in order for a solution to Problem 10 to be much good, it should probably measure growth rates using some alternative number system to the real numbers. Essentially, this is because the real numbers have the Archimedean property (the property that for any positive real $r > 0$ and any real y, there is some $n \in \mathbb{N}$ such that $nr > y$), a constraint which does not apply to function growth rates.

Definition 17 *Let $\mathbb{N}^{\mathbb{N}}$ be the set of all functions $\mathbb{N} \to \mathbb{N}$. A* well-behaved real-measure of $\mathbb{N}^{\mathbb{N}}$ *is a function $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{R}$ satisfying the following requirements.*

1. *(Monotonicity) For each $f, g : \mathbb{N} \to \mathbb{N}$, if $f \succ g$, then $F(f) > F(g)$.*
2. *(Nontriviality) For each $r \in \mathbb{R}$, there is some $f : \mathbb{N} \to \mathbb{N}$ such that $F(f) > r$.*

Theorem 18 *There is no well-behaved real-measure of $\mathbb{N}^{\mathbb{N}}$.*

Proof Assume $F : \mathbb{N}^{\mathbb{N}} \to \mathbb{R}$ is a well-behaved real-measure of $\mathbb{N}^{\mathbb{N}}$. By Nontriviality, there are $f_0, f_1, \ldots : \mathbb{N} \to \mathbb{N}$ such that each $F(f_n) > n$. Define $g : \mathbb{N} \to \mathbb{N}$ by $g(n) = f_0(n) + \cdots + f_n(n)$. Clearly, for every $n \in \mathbb{N}$, $g \succ f_n$. By the Archimedean property of the real numbers, there is some $n \in \mathbb{N}$ such that $n > F(g)$. By Monotonicity, $F(g) > F(f_n)$, but by choice of f_n, $F(f_n) > n > F(g)$, a contradiction. ■

In light of Theorem 18, we are motivated to investigate solutions to Problem 10 using alternatives to the real numbers, which will yield ASPI measures (or taxonomies) in terms of those alternatives.

An informal argument could be made that real numbers might be inadequate for measuring AGI intelligence in general[6]. It at least seems plausible that there are AGIs $X_1, X_2, \ldots$ such that each X_{i+1} is significantly more intelligent than X_i, and another AGI Y such that Y is more intelligent than each X_i. At least, if this is not true, it is not *obvious* that it is not true, and it seems like it would be nontrivial to argue that it is not true[7]. Now, if "significantly more intelligent" implies "at least +1 more intelligent", then it follows that the intelligence levels of Y and of $X_1, X_2, \ldots$ could not all be real numbers, or else one of the X_i would necessarily be more intelligent than Y.

If, as the above argument suggests, the real numbers might potentially be too constrained to perfectly measure intelligence, what next? How could we measure intelligence other than by real numbers? A key motivation for the measures and taxonomies we will come up with below is to provide examples of intelligence measurement using alternative number systems. It is for this reason that we do not, in this paper, consider variations on Hibbard's intelligence measure that arise from simply replacing Liu's enumeration of the primitive recursive functions with various other $\mathbb{N}$-indexed lists of functions (for example, the list of all total computable functions).

6. This argument was pointed out by Alexander (2019b), and by Alexander (2020b) again, the latter amidst a wider discussion of Archimedean and non-Archimedean measures.
7. To do so would require arguing that $\forall X_1, X_2, \ldots$, if each X_{i+1} is significantly more intelligent than X_i, then $\forall Y$, $\exists i$ such that Y is at most as intelligent as X_i.

4. Big-O and Big-Θ intelligence

One of the most standard solutions to Problem 10 in computer science is to categorize growth rates of arbitrary functions by comparing them to more familiar functions using Big-O notation or Big-Θ notation. Knuth (1976) defines these as follows (we modify the definition slightly because we are only concerned here with functions from $\mathbb{N}$ to $\mathbb{N}$).

Definition 19 *Suppose $f : \mathbb{N} \to \mathbb{N}$. We define the following function-sets.*

- *$O(f(n))$ is the set of all $g : \mathbb{N} \to \mathbb{N}$ such that there is some real $C > 0$ and some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $g(n) \leq Cf(n)$.*
- *$\Theta(f(n))$ is the set of all $g : \mathbb{N} \to \mathbb{N}$ such that there are some real $C > 0$ and $C' > 0$ and some $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $Cf(n) \leq g(n) \leq C'f(n)$.*

Note that Definition 19 does not measure growth rates, but rather categorizes growth rates into Big-O and Big-Θ taxonomies. For example, the same function can be both $O(n^2)$ and $O(n^3)$, the former taxon being nested within the latter.

By Remark 16, Definition 19 yields the following elegant taxonomy of predictor intelligence.

Definition 20 *Suppose p is a predictor, and suppose $f : \mathbb{N} \to \mathbb{N}$.*

- *We say p* has Big-O ASPI measure $O(f(n))$ *if p learns every evader e such that t_e is $O(f(n))$.*
- *We say p* has Big-Θ ASPI measure $\Theta(f(n))$ *if p learns every evader e such that t_e is $\Theta(f(n))$.*

Proposition 21 *For any computable function $f : \mathbb{N} \to \mathbb{N}$, there is a predictor p with Big-O ASPI measure $O(f(n))$ and Big-Θ ASPI measure $\Theta(f(n))$.*

Proof Define $g : \mathbb{N} \to \mathbb{N}$ by $g(n) = nf(n) + 1$, and let p_g be as in the proof of Proposition 8. We claim p_g has Big-O ASPI measure $O(f(n))$. To see this, let e be any evader such that t_e is $O(f(n))$. Thus there is some $C \in \mathbb{R}$ such that for all-but-finitely-many $n \in \mathbb{N}$, $t_e(n) \leq Cf(n)$. It follows that $g \succ t_e$. By the Claim in the proof of Proposition 8, for all-but-finitely-many $n \in \mathbb{N}$, if $g(n) > t_e(n)$ then $x_{n+1} = y_{n+1}$, where $(x_1, y_1, \ldots)$ is the result of p_g playing against e. So in all, with only finitely many exceptions, each $x_{n+1} = y_{n+1}$, as desired. The proof that p_g has Big-Θ ASPI measure $\Theta(f(n))$ is similar. ■

5. ASPI measures based on majorization hierarchies

Majorization hierarchies (Weiermann, 2002) provide ordinal-number-valued measures for the growth rates of certain functions. A majorization hierarchy depends on many infinite-dimensional parameters. We will describe two majorization hierarchies up to the ordinal ϵ_0, using standard choices for the parameters, and the ASPI measures which they produce.

Definition 22 *(Classification of ordinal numbers) Ordinal numbers are divided into three types:*

1. *Zero: The ordinal* 0*.*
2. *Successor ordinals: Ordinals of the form* $\alpha + 1$ *for some ordinal* α*.*
3. *Limit ordinals: Ordinals which are not successor ordinals nor* 0*.*

For example, the smallest infinite ordinal, ω, is a limit ordinal. It is not zero (because zero is finite), nor can it be a successor ordinal, because if it were a successor ordinal, say, $\alpha+1$, then α would be finite (since ω is the *smallest* infinite ordinal), but then $\alpha+1$ would be finite as well.

Ordinal numbers have an arithmetical structure: two ordinals α and β have a sum $\alpha+\beta$, a product $\alpha \cdot \beta$, and a power α^β. It would be beyond the scope of this paper to give the full definition of these operations. We will only remark that some care is needed because although ordinal arithmetic is associative—e.g., $(\alpha+\beta)+\gamma = \alpha+(\beta+\gamma)$, and similarly for multiplication—it is not generally commutative: $\alpha + \beta$ is not always equal to $\beta + \alpha$, and $\alpha \cdot \beta$ is not always equal to $\beta \cdot \alpha$. For this reason, one often sees products like $\alpha \cdot 2$, which are not necessarily equivalent to the more familiar $2 \cdot \alpha$.

The ordinal ϵ_0 is the smallest ordinal bigger than the ordinals $\omega, \omega^\omega, \omega^{\omega^\omega}, \ldots$. It satisfies the equation $\epsilon_0 = \omega^{\epsilon_0}$ and can be intuitively thought of as

$$\epsilon_0 = \omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}}.$$

Ordinals below ϵ_0 include such ordinals as ω, $\omega^{\omega+1} + \omega^\omega + \omega^5 + 3$,

$$\omega^{\omega^{\omega^\omega}} + \omega^{\omega^{\omega^\omega} + \omega^{\omega \cdot 2 + 1} + \omega^4 + 3} + \omega^{\omega^5 + \omega^3} + \omega^8 + 1,$$

and so on. Any ordinal below ϵ_0 can be uniquely written in the form

$$\omega^{\alpha_1} + \omega^{\alpha_2} + \cdots + \omega^{\alpha_k}$$

where $\alpha_1 \geq \cdots \geq \alpha_k$ are smaller ordinals below ϵ_0—this form for an ordinal below ϵ_0 is called its *Cantor normal form*. For example, the Cantor normal form for $\omega^{\omega \cdot 2} \cdot 2 + \omega \cdot 3 + 2$ is

$$\omega^{\omega \cdot 2} \cdot 2 + \omega \cdot 3 + 2 = \omega^{\omega \cdot 2} + \omega^{\omega \cdot 2} + \omega^1 + \omega^1 + \omega^1 + \omega^0 + \omega^0.$$

Definition 23 *(Standard fundamental sequences for limit ordinals* $\leq \epsilon_0$*) Suppose* λ *is a limit ordinal* $\leq \epsilon_0$*. We define a* fundamental sequence for λ*, written* $(\lambda[0], \lambda[1], \lambda[2], \ldots)$*, inductively as follows.*

- *If* $\lambda = \epsilon_0$*, then* $\lambda[0] = \omega$*,* $\lambda[1] = \omega^\omega$*,* $\lambda[2] = \omega^{\omega^\omega}$*, and so on.*
- *If* λ *has Cantor normal form* $\omega^{\alpha_1} + \cdots + \omega^{\alpha_k}$ *where* $k > 1$*, then each*

$$\lambda[i] = \omega^{\alpha_1} + \cdots + \omega^{\alpha_{k-1}} + (\omega^{\alpha_k}[i]).$$

- *If λ has Cantor normal form $\omega^{\alpha+1}$, then each $\lambda[i] = \omega^\alpha \cdot i$.*
- *If λ has Cantor normal form ω^{λ_0} where λ_0 is a limit ordinal, then each $\lambda[i] = \omega^{\lambda_0[i]}$.*

Example 24 *(Fundamental sequence examples)*

- *The fundamental sequence for $\lambda = \omega = \omega^{0+1}$ is $\omega^0 \cdot 0, \omega^0 \cdot 1, \omega^0 \cdot 2, \ldots$, i.e., $0, 1, 2, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^5$ is $0, \omega^4, \omega^4 \cdot 2, \omega^4 \cdot 3, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^\omega$ is $\omega^0, \omega^1, \omega^2, \ldots$.*
- *The fundamental sequence for $\lambda = \omega^\omega + \omega$ is $\omega^\omega + 0, \omega^\omega + 1, \omega^\omega + 2, \ldots$.*

Definition 25 *(The standard slow-growing hierarchy up to ϵ_0) We define functions $g_\beta : \mathbb{N} \to \mathbb{N}$ (for all ordinals $\beta \leq \epsilon_0$) by transfinite induction as follows.*

- $g_0(n) = 0$.
- $g_{\alpha+1}(n) = g_\alpha(n) + 1$ *if* $\alpha + 1 \leq \epsilon_0$.
- $g_\lambda(n) = g_{\lambda[n]}(n)$ *if* $\lambda \leq \epsilon_0$ *is a limit ordinal.*

Here are some early levels in the slow-growing hierarchy, spelled out in detail.

Example 26 *(Early examples of functions in the slow-growing hierarchy)*

1. $g_1(n) = g_{0+1}(n) = g_0(n) + 1 = 0 + 1 = 1$.
2. $g_2(n) = g_{1+1}(n) = g_1(n) + 1 = 1 + 1 = 2$.
3. *More generally, for all* $m \in \mathbb{N}$, $g_m(n) = m$.
4. $g_\omega(n) = g_{\omega[n]}(n) = g_n(n) = n$.
5. $g_{\omega+1}(n) = g_\omega(n) + 1 = n + 1$.
6. *More generally, for all* $m \in \mathbb{N}$, $g_{\omega+m}(n) = n + m$.
7. $g_{\omega\cdot 2}(n) = g_{(\omega\cdot 2)[n]}(n) = g_{\omega+n}(n) = n + n = n \cdot 2$.

Following Example 26, the reader should be able to fill in the details in the following example.

Example 27 *(More examples from the slow-growing hierarchy)*

1. $g_{\omega^2}(n) = n^2$.
2. $g_{\omega^3}(n) = n^3$.
3. $g_{\omega^\omega}(n) = n^n$.
4. $g_{\omega^{\omega\cdot 3+1}+\omega+5}(n) = n^{3n+1} + n + 5$.

5. $g_{\omega^{\omega^\omega}}(n) = n^{n^n}$.

What about g_{ϵ_0}? Thinking of ϵ_0 as

$$\omega^{\omega^{\omega^{\cdot^{\cdot^{\cdot}}}}},$$

one might expect $g_{\epsilon_0}(n)$ to be

$$n^{n^{n^{\cdot^{\cdot^{\cdot}}}}},$$

but such an infinite tower of natural number exponents makes no sense if $n > 1$. Instead, the answer defies familiar mathematical notation.

Example 28 *(Level ϵ_0 in the slow-growing hierarchy) The values of g_{ϵ_0} are as follows:*

- $g_{\epsilon_0}(0) = 0$.
- $g_{\epsilon_0}(1) = 1^1$.
- $g_{\epsilon_0}(2) = 2^{2^2}$.
- $g_{\epsilon_0}(3) = 3^{3^{3^3}}$.
- *And so on.*

Examples 26–28 illustrate how the slow-growing hierarchy systematically provides a family of reference functions against which any particular function can be compared. This yields a solution to Problem 10: we can declare the growth rate of an arbitrary function $f : \mathbb{N} \to \mathbb{N}$ to be the smallest ordinal $\beta < \epsilon_0$ such that $g_\beta \succ f$ (or ∞ if there is no such β). For any bounded set S of ordinals, there is a canonical upper bound for S, namely, the supremum of S. Thus we obtain an ASPI measure (not just a taxonomy).

Definition 29 *If p is a predictor, the* ASPI measure of p given by the standard slow-growing hierarchy up to ϵ_0 *is defined to be the supremum of S (or ∞ if $\epsilon_0 \in S$), where S is the set of all ordinals $\alpha \leq \epsilon_0$ such that the following condition holds: for every evader e, if $g_\alpha \succ t_e$, then p learns e.*

In Definition 25, in the successor ordinal case, we chose to define $g_{\alpha+1}(n) = g_\alpha(n) + 1$. The resulting majorization hierarchy is referred to as *slow-growing* because in some sense this makes $g_{\alpha+1}$ just barely faster-growing than g_α. Different definitions of $g_{\alpha+1}$ would yield different majorization hierarchies, such as the following.

Definition 30 *(The standard fast-growing hierarchy up to ϵ_0, also known as the Wainer hierarchy) We define functions $h_\beta : \mathbb{N} \to \mathbb{N}$ (for all ordinals $\beta \leq \epsilon_0$) by transfinite induction as follows.*

- $h_0(n) = n + 1$.

- $h_{\alpha+1}(n) = h_\alpha^n(n)$, *where* h_α^n *is the nth iterate of* h_α *(so* $h_\alpha^1(x) = h_\alpha(x)$, $h_\alpha^2(x) = h_\alpha(h_\alpha(x))$, $h_\alpha^3(x) = h_\alpha(h_\alpha(h_\alpha(x)))$, *and so on).*
- $h_\lambda(n) = h_{\lambda[n]}(n)$ *if* λ *is a limit ordinal* $\leq \epsilon_0$.

The functions in the fast-growing hierarchy grow quickly as α grows. It can be shown (Wainer and Buchholz, 1987) that for every computable function f whose totality can be proven from the axioms of Peano arithmetic, there is some $\alpha < \epsilon_0$ such that $h_\alpha \succ f$.

Definition 31 *If* p *is a predictor, the* ASPI measure of p given by the standard fast-growing hierarchy up to ϵ_0 *is defined to be the supremum of* S *(or* ∞ *if* $\epsilon_0 \in S$*), where* S *is the set of all ordinals* $\alpha \leq \epsilon_0$ *such that the following condition holds: for every predictor* e*, if* $h_\alpha \succ t_e$*, then* p *learns* e.

Proposition 32 *For each* $\alpha < \epsilon_0$*, there is a predictor* p *(resp.* q*) whose ASPI measure given by the standard slow-growing (resp. fast-growing) hierarchy up to* ϵ_0 *is* $\geq \alpha$.

Proof Similar to the proof of Proposition 8. ■

Between Definitions 29 and 31, the former offers a finer granularity intelligence measure for the predictors to which it assigns non-∞ intelligence, but the latter assigns non-∞ intelligence to more intelligent predictors.

Definitions 25 and 30 are only two examples of majorization hierarchies. Both the slow- and fast-growing hierarchies can be extended by extending the fundamental sequences of Definition 23 to larger ordinals[8], however, the larger the ordinals become, the more difficult it is to do this, and especially the less clear it is how to do it in any sort of canonical way. There are also other choices for how to proceed at successor ordinal stages besides $g_{\alpha+1}(n) = g_\alpha(n) + 1$ or $h_{\alpha+1}(n) = h_\alpha^n(n)$—for example, one of the oldest majorization hierarchies is the Hardy hierarchy (Hardy, 1904), where $H_{\alpha+1}(n) = H_\alpha(n+1)$. And even for ordinals up to ϵ_0, there are other ways to choose fundamental sequences besides how we defined them in Definition 23—choosing non-canonical fundamental sequences can drastically alter the resulting majorization hierarchy (Weiermann, 1997). All these different majorization hierarchies yield different ASPI measures.

5.1 A remark about ASPI measures and AGI intelligence

All the ASPI measures and taxonomies we have defined so far double as indirect intelligence measures and taxonomies for an AGI, by the argument we made in Subsection 2.1.

For a given AGI X, a priori, we cannot say much about the predictor which X would act as if X were commanded to act as a predictor. But there is one particularly elegant and parsimonious strategy which X might use, a *brute force strategy*, namely:

- Enumerate all the computable functions f which X knows to be total, and for each one, attempt to predict the evader e by assuming that the evader's runtime t_e satisfies

8. Remarkably, the slow-growing hierarchy eventually catches up with the fast-growing hierarchy if both hierarchies are extended to sufficiently large ordinals (Wainer, 1989; Girard, 1981), a beautiful illustration of how counter-intuitive large ordinal numbers can be.

$f \succ t_e$. If the evader proves not to be so majorized (by differing from every computable function whose runtime is so majorized), then move on to the next known total function f, and continue the process.

We do not know for certain which predictor X would imitate when commanded to act as a predictor, but it seems plausible that X would use this brute force strategy or something equivalent.

For an AGI X who uses the above brute force strategy, ASPI measures of X's intelligence would be determined by X's knowledge, namely, by the runtime complexity of the computable functions which X knows to be total. Furthermore, the most natural way for X to know totality of functions with large runtime complexity, is for X to know fundamental sequences for large ordinal numbers, and produce said functions by means of majorization hierarchies[9]. This suggests a connection between

1. ASPI measures like that of Definition 31, and
2. intelligence measures based on which ordinals the AGI knows (Alexander, 2019b).

Indeed, Alexander (2020a) has argued that the task of notating large ordinals is one which spans the entire range of intelligence. This is reminiscent of Chaitin's proposal to use ordinal notation as a goal intended to facilitate evolution—"and the larger the ordinal, the fitter the organism" (Chaitin, 2011)—and Good's observation (Good, 1969) that iterated Lucas-Penrose contests boil down to contests to name the larger ordinal.

6. Hyperreal numbers and surreal numbers

In this section, we will exhibit an abstract ASPI taxonomy based on hyperreal numbers and an abstract ASPI measure based on surreal numbers. We do not assume previous familiarity with either of these number systems.

6.1 The hyperreal ASPI taxonomy

In this subsection, we will begin by considering growth rate *comparison*, which is a strictly simpler problem than growth rate *measurement* (our proposed solution will then lead to a numerical growth rate measure anyway). Given two functions f and g, does f outgrow g or does f not outgrow g? We would like to say that f outgrows g if and only if $f(n) > g(n)$ for "a majority of" $n \in \mathbb{N}$, but it is not clear what "majority" should mean. Certainly if $f(n) > g(n)$ for all but finitely many $n \in \mathbb{N}$, it should be safe to say f outgrows g, and if $f(n) \leq g(n)$ for all but finitely many $n \in \mathbb{N}$, it should be safe to say f does not outgrow g. But what if there are infinitely many $n \in \mathbb{N}$ such that $f(n) > g(n)$, and infinitely many $n \in \mathbb{N}$ such that $f(n) \leq g(n)$?

Adapting the key insight from Alexander (2019a), consider each $n \in \mathbb{N}$ to be a voter in an election to determine whether or not f outgrows g. Each n votes based on whether or not $f(n) > g(n)$. For example, 532 is a voter in this election. If $f(532) > g(532)$, then 532

9. It may be possible for an AGI to be contrived to know totality of functions that are larger than the functions produced by majorization hierarchies up to ordinals the AGI knows about, but we conjecture that that is not the case for AGIs not so deliberately contrived.

casts her vote for "f outgrows g"; otherwise, 532 casts her vote for "f does not outgrow g". This reduces the outgrowth problem to an election decision problem: f shall be considered to outgrow g if and only if "f outgrows g" gets a winning bloc of votes.

We need to decide what it means for a set $N \subseteq \mathbb{N}$ to constitute a winning bloc of votes. We reason as follows.

- $\emptyset$ should not be a winning bloc: if no-one votes for you, you lose.
- If N_1 is a winning bloc and $N_1 \subseteq N_2$, then N_2 should be a winning bloc: if you were already winning, and additional voters switch their votes to you, you should still win.
- For any $N \subseteq \mathbb{N}$, either N should be a winning bloc, or its complement $N^c = \mathbb{N} \backslash N$ should be a winning bloc: however the election goes, either you win or your opponent wins.
- We should insist on the outgrowth relation being transitive: if f outgrows g and g outgrows h, then f should outgrow h. Suppose that

$$\begin{aligned} N_{fg} &= \{n \in \mathbb{N} : f(n) > g(n)\}, \\ N_{gh} &= \{n \in \mathbb{N} : g(n) > h(n)\}, \text{ and} \\ N_{fh} &= \{n \in \mathbb{N} : f(n) > h(n)\}. \end{aligned}$$

  Clearly $N_{fg} \cap N_{gh} \subseteq N_{fh}$ but, a priori, we cannot say more: one can find functions f, g, h such that $N_{fg} \cap N_{gh} = N_{fh}$. Thus, in order to ensure transitivity of the outgrowth relation, we should insist on the following requirement. Whenever N_1 and N_2 are winning blocs, then $N_1 \cap N_2$ should be a winning bloc.
- We could trivially satisfy the above requirements, namely: we could choose some $n_0 \in \mathbb{N}$ and declare that $N \subseteq \mathbb{N}$ is a winning bloc if and only if $n_0 \in N$. In electoral terms, this would amount to making n_0 a dictator, whose vote decides the election regardless how anyone else votes. In terms of the outgrowth relation, this would amount to declaring that f outgrows g if and only if $f(n_0) > g(n_0)$. That would be a poor method of comparing growth rates. Thus, we should insist that $\{n_0\}$ is not a winning bloc for any $n_0 \in \mathbb{N}$.

Is it possible to satisfy all the above requirements, or are they too demanding? It turns out it is possible. In fact, the above requirements are exactly the requirements of a *free ultrafilter*, an important device from mathematical logic.

Definition 33 *An* ultrafilter on $\mathbb{N}$ *(or more simply an* ultrafilter*) is a set $\mathcal{U}$ of subsets of $\mathbb{N}$ such that:*

1. *$\emptyset \notin \mathcal{U}$.*
2. *For every $N_1 \in \mathcal{U}$, for every $N_2 \subseteq \mathbb{N}$, if $N_1 \subseteq N_2$, then $N_2 \in \mathcal{U}$.*
3. *For every $N \subseteq \mathbb{N}$, either $N \in \mathcal{U}$ or $N^c \in \mathcal{U}$.*
4. *($\cap$-closure) For every $N_1, N_2 \in \mathcal{U}$, $N_1 \cap N_2 \in \mathcal{U}$.*

An ultrafilter is free *if it does not contain any singleton* $\{n_0\}$ *(*$n_0 \in \mathbb{N}$*).*

Clearly a free ultrafilter is exactly a notion of winning blocs meeting all our requirements. The following theorem is well-known in mathematical logic, and we state it without proof.

Theorem 34 *Free ultrafilters exist.*

Theorem 34 is profound because it is counter-intuitive that there should be a non-dictatorial method of determining election winners satisfying $\cap$-closure. To see how counter-intuitive $\cap$-closure is, suppose that in 2021 the Dog party wins the presidency and in 2022 the Cat party wins the presidency (with the same voters every year and no other parties). Call a voter a "Dog-to-Cat switcher" if they vote Dog in 2021 and Cat in 2022. The $\cap$-closure property says in order to win in 2023, it would be enough to get just the Dog-to-Cat switchers' votes *and no others*. For more on infinite-voter elections and free ultrafilters, and especially their interplay with Arrow's impossibility theorem, see Kirman and Sondermann (1972).

For the remainder of the section, let $\mathcal{U}$ be a free ultrafilter. Unfortunately, logicians have shown that, though free ultrafilters exist, it is impossible to concretely exhibit one. More precisely, all known proofs of Theorem 34 are non-constructive (using non-constructive set-theoretic axioms such as the Axiom of Choice) and logicians have proven that Theorem 34 cannot be proved constructively.

Definition 35 *Suppose* $f, g : \mathbb{N} \to \mathbb{R}$*. We say* $f >_{\mathcal{U}} g$ *if*

$$\{n \in \mathbb{N} : f(n) > g(n)\} \in \mathcal{U}.$$

In other words: if $\mathcal{U}$ *is thought of as a black box deciding which subsets of* $\mathbb{N}$ *are winning blocs, then* $f >_{\mathcal{U}} g$ *if and only if "*f *outgrows* g*" wins the election when each* $n \in \mathbb{N}$ *votes for "*f *outgrows* g*" or "*f *does not outgrow* g*" depending whether* $f(n) > g(n)$ *or* $f(n) \leq g(n)$ *respectively.*

Lemma 36 $>_{\mathcal{U}}$ *is transitive.*

Proof Suppose $f, g, h : \mathbb{N} \to \mathbb{R}$ are such that $f >_{\mathcal{U}} g$ and $g >_{\mathcal{U}} h$, we must show $f >_{\mathcal{U}} h$. Let

$$\begin{aligned} N_{fg} &= \{n \in \mathbb{N} : f(n) > g(n)\}, \\ N_{gh} &= \{n \in \mathbb{N} : g(n) > h(n)\}, \text{ and} \\ N_{fh} &= \{n \in \mathbb{N} : f(n) > h(n)\}. \end{aligned}$$

Since $f >_{\mathcal{U}} g$, $N_{fg} \in \mathcal{U}$. Since $g >_{\mathcal{U}} h$, $N_{gh} \in \mathcal{U}$. By $\cap$-closure, $N_{fg} \cap N_{gh} \in \mathcal{U}$. Clearly $N_{fg} \cap N_{gh} \subseteq N_{fh}$, so, by (2) of Definition 33, $N_{fh} \in \mathcal{U}$, that is, $f >_{\mathcal{U}} h$. ■

We will now explain how Definition 35 leads to a numerical growth rate measure and, in turn, an ASPI taxonomy. We will show that by coming up with Definition 35, we have actually done much of the work of the so-called ultrapower construction of the hyperreal number system, studied in the field of non-standard analysis (Robinson, 1974; Goldblatt, 2012).

Definition 37 *(Compare Definition 35) Suppose $f, g : \mathbb{N} \to \mathbb{R}$. We say $f \equiv g$ if*

$$\{n \in \mathbb{N} : f(n) = g(n)\} \in \mathcal{U}.$$

In other words, $f \equiv g$ if "$f = g$" wins the election (as decided by $\mathcal{U}$) when each $n \in \mathbb{N}$ votes for "$f = g$" or "$f \neq g$" depending whether $f(n) = g(n)$ or $f(n) \neq g(n)$ respectively.

Lemma 38 *The relation $\equiv$ (from Definition 37) is an equivalence relation.*

Proof Symmetry and reflexivity are trivial. The proof of transitivity is similar to Lemma 36. ■

Definition 39 *The* hyperreal numbers*, written ${}^*\mathbb{R}$, are the equivalence classes of $\equiv$. For every $f : \mathbb{N} \to \mathbb{R}$, write $\hat{f}$ for the hyperreal number (i.e., the $\equiv$-equivalence class) containing f. We endow ${}^*\mathbb{R}$ with arithmetic and order as follows (where $f, g : \mathbb{N} \to \mathbb{R}$):*

- *We define addition on ${}^*\mathbb{R}$ by declaring that $\hat{f} + \hat{g} = \hat{h}$ where $h(n) = f(n) + g(n)$.*
- *We define multiplication on ${}^*\mathbb{R}$ by declaring that $\hat{f} \cdot \hat{g} = \hat{h}$ where $h(n) = f(n)g(n)$.*
- *We order ${}^*\mathbb{R}$ by declaring that $\hat{f} > \hat{g}$ if and only if $f >_{\mathcal{U}} g$ (Definition 35).*

The following is well-known and we state it without proof.

Theorem 40 *The addition, multiplication, and ordering in Definition 39 are well-defined, and they make ${}^*\mathbb{R}$ an ordered field.*

With this machinery, we now have a trivial hyperreal solution to Problem 10.

Definition 41 *(The hyperreal solution to Problem 10) For any function $f : \mathbb{N} \to \mathbb{N}$, the* hyperreal growth rate of *f is the hyperreal number $\hat{f}$.*

Because of the non-constructive nature of free ultrafilters, the following notions are even less practical than the measures in the previous sections. However, they could potentially be useful for proving theoretical properties about the intelligence of predictors.

Definition 42 *Suppose p is a predictor and e is an evader. Let $(x_1, y_1, x_2, y_2, \ldots)$ be the result of p playing against e. We say p $\mathcal{U}$-*learns *e if*

$$\{n \in \mathbb{N} : p(x_1, \ldots, x_n) = e(y_1, \ldots, y_n)\} \in \mathcal{U}$$

(or equivalently: $\{n \in \mathbb{N} : y_{n+1} = x_{n+1}\} \in \mathcal{U}$). In other words, p $\mathcal{U}$-learns e if "p learns e" wins the election (according to $\mathcal{U}$) when every $n \in \mathbb{N}$ votes for "p learns e" or "p does not learn e" depending whether or not $p(x_1, \ldots, x_n) = e(y_1, \ldots, y_n)$.

In the following definition, rather than assigning a particular hyperreal number intelligence to every predictor, rather, we categorize predictors into a taxonomy. This is necessary because there is no way of choosing canonical bounds of bounded sets of hyperreal numbers in general. For lack of a way of choosing a particular bound, we are forced to consider many taxa corresponding to many bounds.

Definition 43 *(The hyperreal ASPI taxonomy) Let p be a predictor and let $\hat{f}$ be a hyperreal number. We say that p* has hyperreal ASPI intelligence at least *$\hat{f}$ if and only if the following condition holds: for every evader e, if the hyperreal growth rate of t_e is $< \hat{f}$, then p $\mathcal{U}$-learns e.*

Now we would like to state an analog of Proposition 8 for hyperreal ASPI intelligence, but before we can do that, we need to state the following lemma. This lemma is well-known so we state it without proof.

Lemma 44 *For any $N \in \mathcal{U}$, for any finite $N_0 \subseteq \mathbb{N}$, the difference $N \backslash N_0$ is in $\mathcal{U}$.*

Proposition 45 *For any computable function $f : \mathbb{N} \to \mathbb{N}$, there is a predictor which has hyperreal ASPI intelligence at least $\hat{f}$.*

Proof Let p_f be as in the proof of Proposition 8. We claim p_f has hyperreal ASPI intelligence at least $\hat{f}$. To see this, assume e is an evader such that the hyperreal growth rate of t_e is $< \hat{f}$, we will show p_f $\mathcal{U}$-learns e.

Let $(x_1, y_1, x_2, y_2, \ldots)$ be the result of p_f playing against e. Let

$$N_1 = \{n \in \mathbb{N} : f(n) > t_e(n)\},$$
$$N_2 = \{n \in N_1 : p_f(x_1, \ldots, x_n) = e(y_1, \ldots, y_n)\},$$
$$N_3 = \{n \in \mathbb{N} : p_f(x_1, \ldots, x_n) = e(y_1, \ldots, y_n)\}.$$

By the Claim in the proof of Proposition 8, for all-but-finitely-many $n \in \mathbb{N}$, if $f(n) > t_e(n)$, then $x_{n+1} = y_{n+1}$, i.e., $p_f(x_1, \ldots, x_n) = e(y_1, \ldots, y_n)$. In other words, $N_2 = N_1 \backslash N_0$ for some finite $N_0 \subseteq \mathbb{N}$.

Since the hyperreal growth rate of t_e is $< \hat{f}$, $N_1 \in \mathcal{U}$. Since $N_2 = N_1 \backslash N_0$, by Lemma 44 we have $N_2 \in \mathcal{U}$. Since $N_2 \subseteq N_3$, it follows that $N_3 \in \mathcal{U}$, that is, p_f $\mathcal{U}$-learns e, as desired. ■

Whereas we showed in Proposition 9 that the original Hibbard measure is already non-computable, we conjecture that the hyperreal ASPI taxonomy is even harder (in a computability theoretical sense).

We mention in passing that there are other ways to approach non-standard analysis (Hrbacek and Katz, 2020), where the free ultrafilter dependency is replaced by a dependency on a model of a certain set of axioms.

6.2 The surreal ASPI measure

In Definition 43, we had to content ourselves with an ASPI taxonomy rather than an ASPI measure, because there is no canonical way of choosing a preferred bound for a bounded set of hyperreals. In this section, we will consider another number system, the surreal numbers (Conway, 2000; Knuth, 1974; Ehrlich, 2012), in which the hyperreal numbers can be embedded. By embedding the hyperreals within the surreals, our hyperreal solution to Problem 10 yields a surreal solution to Problem 10, via the embedding. Unlike the hyperreal numbers, the surreal numbers *do* admit a canonical way of choosing preferred bounds for

bounded sets. Thus, our surreal solution to Problem 10 will yield an ASPI measure, not just a taxonomy.

Formally defining the surreal numbers involves nuances touching the foundations of mathematics, so we will only sketch the definition here. Suppose we take the following as the guiding principle in creating a number system:

- For every set L of lower bounds in our number system, and every set R of upper bounds in our number system, with $L < R$ (by which we mean $\ell < r$ for all $\ell \in L, r \in R$), we want there to exist a canonical number $L < x < R$ (i.e., a canonical number x such that $\ell < x < r$ for all $\ell \in L, r \in R$).

To simplify things, since this guiding principle only requires the existence of these canonical numbers, we can assume no other numbers exist, only such canonical numbers. Thus, every number in the resulting system can, recursively, be *identified* with a pair (L, R) of sets of numbers, $L < R$. Notationally, we use $\{L|R\}$ as a name for the canonical number $L < \{L|R\} < R$ between L and R, whenever $L < R$ are sets of numbers. There are two tricky issues with this idea:

1. A number in this number system we are building might have multiple names, so we cannot simply take the numbers to be their names. Instead, it is necessary to take the numbers to be equivalence classes of names. What should it mean for two names to be equivalent?

2. If $\{x_L|x_R\}$ and $\{y_L|y_R\}$ are (names of) two numbers x, y (respectively) in this number system we are building (so $x_L < x_R$ and $y_L < y_R$ are sets of numbers in said number system), what does it mean to say $x < y$? This question seems somewhat circular, because in order for $\{x_L|x_R\}$ to be a valid name in the first place already requires that $x_L < x_R$ (i.e., that $\ell < r$ whenever $\ell \in x_L, r \in x_R$).

To answer issue 2, we ask ourselves: what does it mean for the canonical number x between x_L and x_R to be less than the canonical number y between y_L and y_R? With some thought, it seems a natural way to (recursively) answer this question is to declare $x < y$ if and only if one of the following holds:

- There is some $y' \in y_L$ such that $x \leq y'$, or
- There is some $x' \in x_R$ such that $x' \leq y$.

Having answered (2), we can answer (1) by declaring that two names $\{x_L|x_R\}$ and $\{y_L|y_R\}$ (of x and y respectively) are equivalent if $x \not< y$ and $y \not< x$.

Carrying out the above definition in full formality is tricky, because the $<$ relation and the equivalence relation have to be defined by simultaneous recursion in terms of each other. To give the formal definition here would be beyond the scope of this paper[10]. The equivalence classes of the above equivalence relation are called *surreal numbers*. It is possible to define addition and multiplication on them in such a way that, along with the ordering $<$ already defined, the surreal numbers satisfy the axioms of an ordered field.

10. Technically, $<$ and the equivalence relation are not actually relations at all, because their universes are not sets. For this and other reasons, the formal construction of the surreals is usually carried out in transfinitely many stages, such that at any particular stage, the surreals constructed so far form a set.

Example 46 *(Examples of surreal numbers)*

- *Taking $L = R = \emptyset$ yields the surreal number named $\{\emptyset|\emptyset\}$, usually abbreviated $\{|\}$. It can be shown this is the surreal 0 (i.e., the unique surreal additive identity).*
- *Let $L = \{0\}$, $R = \emptyset$, where 0 is as above. This yields the surreal named $\{\{0\}|\}$. It can be shown this is the surreal 1 (i.e., the unique multiplicative identity).*
- *Let $L = \{1\}$, $R = \emptyset$, where 1 is as above. This yields the surreal number named $\{\{1\}|\}$. It can be shown this is 2 (i.e., $1 + 1$).*
- *In the same way, one can obtain surreal numbers $3, 4, 5, \ldots$.*
- *Let $L = \{0, 1, 2, \ldots\}$ (as above), $R = \emptyset$. This yields the surreal named $\{\{0, 1, 2, \ldots\}|\}$. This surreal represents the smallest infinite ordinal number, ω, considered as a surreal.*
- *With 0 and 1 as above, $\{\{0\}|\{1\}\}$ names a surreal strictly between 0 and 1. It can be shown to be $\frac{1}{2}$ (i.e., that its product with 2 is 1).*
- *With 0 and $\frac{1}{2}$ as above, $\{\{0\}|\{\frac{1}{2}\}\}$ names a surreal strictly between 0 and $\frac{1}{2}$. It can be shown to be $\frac{1}{4}$.*
- *Similarly, one can construct surreals $\frac{1}{8}, \frac{1}{16}, \ldots$.*
- *With 0 and $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$ as above, $\{\{0\}|\{\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots\}\}$ names a surreal larger than 0 but smaller than every $\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \ldots$. This shows that the surreals include infinitesimals. This particular infinitesimal can be shown to be $1/\omega$, in the sense that when multiplied by ω, the result is 1.*

The reader might object that if we let L be the set of all surreals, then $\{L|\}$ seems to name a surreal larger than all surreals, which would be absurd. This paradox is avoided because in fact the class of all surreals is *not* a set, but a proper class[11].

It can be shown that for any free ultrafilter $\mathcal{U}$, the hyperreals constructed using $\mathcal{U}$ can be embedded into the surreals. This allows us to transform our hyperreal solution to Problem 10 into a surreal solution. Unfortunately, there are many different ways to embed the hyperreals into the surreals, none of them canonical, so while our hyperreal solution already depends arbitrarily on a free ultrafiler, our surreal solution further depends on an arbitrary embedding.

Definition 47 *(The surreal solution to Problem 10) Suppose $\mathcal{U}$ is a free ultrafilter and ${}^*\mathbb{R}$ is the corresponding hyperreal number system (Definition 39). Let ι be an embedding of ${}^*\mathbb{R}$ into the surreals. For any function $f : \mathbb{N} \to \mathbb{N}$, the* surreal growth rate of f *(given by $\mathcal{U}$ and ι) is $\iota(\hat{f})$ where $\hat{f}$ is the hyperreal growth rate of f (Definition 41).*

The payoff of doing this is that there is a canonical way to pick a particular bound of a bounded set of surreals, so the surreal solution to Problem 10 provides an ASPI measure, not just an ASPI taxonomy.

11. It can be shown that the ordinal numbers can be embedded in the surreals, and so the non-sethood of the class of surreals is strictly weaker than the non-sethood of the class of ordinals. The latter non-sethood is referred to as the Burali-Forti paradox.

Definition 48 *(The surreal ASPI measure) Let $\mathcal{U}$ be a free ultrafilter, let ${}^*\mathbb{R}$ be the corresponding hyperreal number system, and let ι be an embedding of ${}^*\mathbb{R}$ into the surreals. Let $\iota({}^*\mathbb{R})$ be the range of ι. For every predictor p, the* surreal ASPI measure *of p (given by $\mathcal{U}$ and ι) is defined to be the surreal with name $\{L|\}$, where L is the set of all surreal numbers $\ell \in \iota({}^*\mathbb{R})$ such that the following condition holds: for every evader e, if the surreal growth rate of t_e is $< \ell$, then p $\mathcal{U}$-learns e.*

Note that in Definition 48, we require $\ell \in \iota({}^*\mathbb{R})$ because otherwise L would not be a set.

7. Pros and cons of different ASPI measures and taxonomies

Here are pros and cons of the ASPI measures and taxonomies which arise from different solutions to the problem (Problem 10) of measuring the growth rate of functions.

- The original Hibbard measure (Definition 7), which arises by measuring growth rate by comparing a function with Liu's enumeration (Liu, 1960) of the primitive recursive functions:
  - Pro: Relatively concrete.
  - Pro: Measures intelligence using a familiar number system (the natural numbers).
  - Con: The numbers which the measure outputs are not very meaningful, in that predictor p having a measure of $+1$ higher than predictor q tells us little about how *much* more computationally complex the evaders which p learns are, versus the evaders which q learns.
  - Con: Only distinguishes sufficiently non-intelligent predictors; all predictors sufficiently intelligent receive measure ∞.
- Big-O/Big-Θ (Definition 20), in which, rather than directly measuring the intelligence of a predictor, instead, we would talk of a predictor's intelligence being $O(f(n))$ or $\Theta(f(n))$ for various functions $f : \mathbb{N} \to \mathbb{N}$:
  - Pro: Nearly perfect granularity (slightly coarser than perfect granularity because of the constants C, C' in Definition 19).
  - Pro: Computer scientists already use Big-O/Big-Θ routinely and are comfortable with them.
  - Con: A non-numerical taxonomy.
- Intelligence based on a majorization hierarchy such as the standard slow- or fast-growing hierarchy up to ϵ_0 (Definitions 29 and 31):
  - Pro: A numerical measure, albeit less granular than the Big-O/Big-Θ taxonomies.
  - Pro: Relatively concrete.
  - Pro: The numbers which the measure outputs are meaningful, in the sense that the degree to which a predictor p is more intelligent than a predictor q is reflected in the degree to which p's intelligence-measure is larger than q's.

  - Con: The numbers which the measure outputs are ordinal numbers, which may be unfamiliar to some users.
  - Con: Only distinguishes sufficiently non-intelligent predictors; for any particular majorization hierarchy, all predictors sufficiently intelligent receive measure ∞.

- Hyperreal intelligence (Definition 43):
  - Pro: A taxonomy like Big-O/Big-Θ, but with the added benefit that the taxons are numerical.
  - Pro: Perfect granularity.
  - Con: Depends on a free ultrafilter (free ultrafilters exist but cannot be concretely exhibited).

- Surreal intelligence (Definition 48):
  - Pro: An actual numerical measure (not just a taxonomy), with perfect granularity.
  - Con: The numbers which the measure outputs are surreal numbers, which are relatively new and thus unfamiliar, and are difficult to work with in practice.
  - Con: Depends on both a free ultrafilter and also an embedding of the resulting hyperreals into the surreals.

8. Conclusion

To summarize:

- Hibbard (2011) proposed an intelligence measure for predictors in games of adversarial sequence prediction.

- We argued that Hibbard's idea actually splits into two orthogonal sub-ideas. First: that intelligence can be measured via the growth-rates of the run-times of evaders that a predictor can learn to predict. Second: that such growth-rates can be measured in one specific way (involving an enumeration of the primitive recursive functions). We argued that there are many other ways to measure growth-rates, and that each method of measuring growth-rates yields a corresponding adversarial sequence prediction intelligence (ASPI) measure or taxonomy.

- We considered several specific ways of measuring growth-rates of functions, and exhibited corresponding ASPI measures and taxonomies. The growth-rate-measuring methods which we considered were: Big-O/Big-Θ notation; majorization hierarchies; hyperreal numbers; and surreal numbers.

- We also discussed how the intelligence of adversarial sequence predictors can be considered as an approximation of the intelligence of idealized AGIs.

Acknowledgments

We acknowledge Bryan Dawson for feedback on Section 6.1. We acknowledge Philip Ehrlich for correcting a mistake. We acknowledge Mikhail Katz and Roman Yampolskiy for providing literature references. We acknowledge the editor and the reviewers for much generous feedback and suggestions.

References

Alexander, S. A. 2019a. Intelligence via ultrafilters: structural properties of some intelligence comparators of deterministic Legg-Hutter agents. *Journal of Artificial General Intelligence* 10(1):24–45.

Alexander, S. A. 2019b. Measuring the intelligence of an idealized mechanical knowing agent. In *CIFMA*.

Alexander, S. A. 2020a. AGI and the Knight-Darwin Law: why idealized AGI reproduction requires collaboration. In *ICAGI*.

Alexander, S. A. 2020b. The Archimedean trap: Why traditional reinforcement learning will probably not yield AGI. *Journal of Artificial General Intelligence* 11(1):70–85.

Bostrom, N. 2003. Ethical issues in advanced artificial intelligence. In Schneider, S., ed., *Science fiction and philosophy: from time travel to superintelligence*. John Wiley and Sons. 277–284.

Chaitin, G. 2011. Metaphysics, Metamathematics and Metabiology. In Hector, Z., ed., *Randomness through computation*. World Scientific.

Conway, J. H. 2000. *On Numbers and Games*. CRC Press, 2nd edition.

Ehrlich, P. 2012. The absolute arithmetic continuum and the unification of all numbers great and small. *Bulletin of Symbolic Logic* 18:1–45.

Girard, J.-Y. 1981. Π^1_2-logic, Part 1: Dilators. *Annals of Mathematical Logic* 21(2-3):75–219.

Goldblatt, R. 2012. *Lectures on the hyperreals: an introduction to nonstandard analysis*. Springer.

Good, I. J. 1969. Gödel's theorem is a red herring. *The British Journal for the Philosophy of Science* 19(4):357–358.

Hardy, G. H. 1904. A theorem concerning the infinite cardinal numbers. *Quarterly Journal of Mathematics* 35:87–94.

Hibbard, B. 2008. Adversarial sequence prediction. In *ICAGI*, 399–403.

Hibbard, B. 2011. Measuring agent intelligence via hierarchies of environments. In *ICAGI*, 303–308.

Hrbacek, K., and Katz, M. G. 2020. Infinitesimal analysis without the axiom of choice. *Preprint*.

Hutter, M. 2004. *Universal artificial intelligence: Sequential decisions based on algorithmic probability*. Springer.

Kirman, A. P., and Sondermann, D. 1972. Arrow's theorem, many agents, and invisible dictators. *Journal of Economic Theory* 5(2):267–277.

Knuth, D. E. 1974. *Surreal numbers: a mathematical novelette*. Addison-Wesley.

Knuth, D. E. 1976. Big Omicron and big Omega and big Theta. *ACM Sigact News* 8(2):18–24.

Legg, S. 2006. Is there an elegant universal theory of prediction? In *International Conference on Algorithmic Learning Theory*, 274–287. Springer.

Liu, S.-C. 1960. An enumeration of the primitive recursive functions without repetition. *Tohoku Mathematical Journal* 12(3):400–402.

Robinson, A. 1974. *Non-standard analysis*. Princeton University Press.

Wainer, S., and Buchholz, W. 1987. Provably computable functions and the fast growing hierarchy. In Simpson, S. G., ed., *Logic and Combinatorics*. AMS.

Wainer, S. 1989. Slow growing versus fast growing. *The Journal of Symbolic Logic* 54(2):608–614.

Wang, P. 2019. On Defining Artificial Intelligence. *Journal of Artificial General Intelligence* 10(2):1–37.

Weiermann, A. 1997. Sometimes slow growing is fast growing. *Annals of Pure and Applied Logic* 90(1-3):91–99.

Weiermann, A. 2002. Slow versus fast growing. *Synthese* 133:13–29.

Yampolskiy, R. V. 2012. AI-complete, AI-hard, or AI-easy–classification of problems in AI. In *The 23rd Midwest Artificial Intelligence and Cognitive Science Conference*.

Yampolskiy, R. V. 2013. Turing test as a defining feature of AI-completeness. In *Artificial intelligence, evolutionary computing and metaheuristics*. Springer. 3–17.

Yampolskiy, R. V. 2020. On Controllability of Artificial Intelligence. *Technical report*.